ITM Web of Conferences **72**, 03008 (2025) https://doi.org/10.1051/itmconf/20257203008
HMMOCS-III 2024

Analysis of the dynamics of changes in factor load models for predicting failures of a spacecraft

Konstantin Kraev[1*], *Pavel* Zaitsev[2], and *Vladimir* Kazakovtsev[1,2]

[1]Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarskii rabochii prospekt, Krasnoyarsk 660037, Russian Federation
[2]Siberian Federal University, 79 Svobodny Prospekt, Krasnoyarsk 660041, Russian Federation

Abstract. The paper considers the theoretical aspects of factor analysis as applied to the problem of predicting the failure-free operation of low-orbit communication satellites. The authors described the nature of the factors affecting the spacecraft during operation. The structure of the methodology for preparing data and creating factor load models is described. A methodology for analyzing the dynamics of changes in factor load models is proposed. The basic concepts and the algorithm are described. A numerical experiment to test the proposed hypothesis and interpreted the key factors influencing the functioning of the spacecraft is described. The factors are distributed into influence groups. As a result of the numerical experiment and the conducted analysis of the dynamics of changes in the factor load models, we made a conclusion on the validity of the proposed hypothesis. The practical applicability of the described approach for predicting failures and malfunctions during the operation of spacecraft is discussed.

1 Introduction

When operating low-orbit satellite systems, the issue of forecasting their further functioning is very relevant. To solve this problem, it is necessary to take into account the influence of external and internal factors affecting the spacecraft during its operation. The methods of factor analysis allow us to present the influence of factors affecting the spacecraft in the form of models of factor loads. A large number of variables related to the available observations are replaced by a smaller number of independent influencing variables called factors. A set of input data is taken as variables. The reduction in the number of variables is based on their classification and determination of the structure of the relationships between them. Due to this reduction, it becomes possible to analyze data on selected factors, the number of which is significantly less than the initial number of interrelated variables. Therefore, factor analysis is used either as a data reduction method or as a classification method. It makes it possible to simplify complex systems, identify key factors and improve data interpretation [1, 2].

The following classification of factors is accepted in technical tasks [3]: constructive, production, operational.

* Corresponding author: levk@bk.ru

We assume that the influence of design and production factors has already been taken into account during the design and manufacture of the spacecraft according to the technical specifications. Therefore, to predict the functioning of the low-orbit communication satellites, it is necessary to conduct a factor analysis of the influence of operating conditions.

The factor load models obtained as a result of the factor analysis can be used to test the following hypothesis: factor models calculated for different periods of operation of a technical object should differ depending on changes in the technical condition of this object. The purpose of the study is to analyze the dynamics of changes in the obtained factor load models using low-orbit communication satellites as an example.

To determine the factors affecting the operability of the spacecraft and to conduct a factor analysis, it is necessary to collect and process data (from external sources, telemetry and ballistic data). Today, there are many technical and mathematical solutions for collecting telemetry data on the state of the spacecraft that are constantly being improved [4-7]. At the same time, the procedures for using these data to conduct a factor analysis in order to predict the further functioning of the spacecraft have not been studied. To solve this problem, it is proposed to analyze the dynamics of changes in the models of spacecraft factor loads. Today, this approach looks promising due to the constant development of Data Science methods that allow analyzing large volumes of data.

2 Methodology for data preparation and creation of factor loading models

The proposed methodology is based on well-known methods of factor analysis for obtaining, processing and converting unstructured data. As a rule, the initial data will have large volumes. The greater the number of input variables on operational impacts on the spacecraft and the number of observations of these impacts, the better the quality of the analysis results will be. At the same time, working with large volumes of data is associated with high complexity. Therefore, the application of Data Science methods will be very relevant in the process of data processing [8].

Let's consider the main stages of the methodology in more detail.

Data collection. It is necessary to collect the data that will be analyzed.

All collected data must be quantitative. If this is not the case, then they must be coded. The number of observations for analysis must be at least 2 times greater than the number of variables. Otherwise, the analysis will be difficult.

The collected data may have different time scales of observations. For example, some data is collected once a day, and other data - 4-5 times a day. To carry out the analysis, it is necessary to reduce all data to a single time scale.

Data preparation. During the collection of information, anomalous data (outliers) may appear, which may be due to faulty measurements, errors in data collection or input, etc. [9]. Outliers can distort the results of the analysis. Today, there are many statistical methods for excluding these data: interquartile methods, graphical methods, clustering methods.

Data processing. The data for analysis usually have different dimensions. In order to bring them to the same dimension and ensure comparability, the source data is usually normalized by introducing a single format. The most common method of data normalization is standardization. It is used when the ranges of data changes from different sources differ significantly. Standardization is a method of changing the ranges of values - scaling. The procedure moves all variables to similar scales by subtracting the mean value of the variable and dividing by the standard deviation (z-standardization):

$$z = \frac{x - \overline{x}}{s}, \quad (1)$$

where $\bar{x}$ is the mean value of the variable, s is the standard deviation.

Construction of the correlation matrix. Finding the correlation between all variables to determine their interrelations with each other. With a small number of variables, a visual analysis of this matrix can be carried out. As the number of variables increases (10 or more), visual analysis will not give positive results. The advisability of conducting factor analysis is determined by the presence of correlations between the variables.

Checking the significance of the initial correlation matrix. There are several criteria for evaluating the original correlation matrix: the Bartlett-Wilks criterion of sphericity and the Kaiser criterion of sampling adequacy.

Criteria for a preliminary assessment of the number of identified factors. Currently, the Kaiser criterion and the Cattell screening criterion ("rocky scree") are used.

Factor identification. Factor analysis methods are used to identify the main factors that explain most of the data variability. Factor analysis methods differ depending on the approaches used to identify the coefficients of the factor values. The following methods can be distinguished: principal component analysis and common factor analysis (maximum likelihood method, least squares method, etc.). The choice of a specific method depends on the available data and the purposes of the analysis. The principal component method is simpler and has proven itself well. This method will be used in factor analysis in the future because it describes the maximum variance of input characteristics.

The Principal Components Analysis method is based on determining the minimum number of factors that make the greatest contribution to data variance. They are called principal components [10].

Factor rotation. Rotation is used to find one of the possible coordinate systems in the factor space that simplifies the factor structure. When rotating the coordinate system formed by the factors, the accuracy of data presentation through the new axes is not lost, but only the ordering of the factors by the magnitude of the variance explained by them changes.

Obtaining new factors will maximize high correlations and minimize low ones. As a result, the factor loadings will be more contrasting, which will facilitate further interpretation of the factors [11-12].

Interpretation of factors. The analysis of the selected factors is carried out in order to understand their meaning and determine how they affect the original variables. The main information for the analysis is the factor loadings. To interpret the factors, it is necessary to give each factor a term or concept. This concept appears on the basis of the analysis of the correlations of the factor with the original variables. Based on the obtained data, various hypotheses can be built.

It should be borne in mind that the interpretation of factors is largely associated with understanding the behavior of the object under study. There may be situations when the result of interpreting the factors will be illogical for understanding. In this case, there is a possibility that during the interpretation, some hidden implicit factor was found that somehow affects the object of study.

As a result of the described stages of the data preparation methodology, models are formed, presented in the form of interpreted matrices of factor loadings (hereinafter - MFL) of the object under study. Subsequent analysis of these models will allow a comparison of MFLs calculated for different periods of the object's life cycle. According to the proposed hypothesis, the matrices should differ depending on the change in the state of this object.

3 Methodology for analyzing the dynamics of changes in factor loading models

We consider N objects which are subject to analysis, divided into three classes and studied in n observation periods:

– Class I - the object is in a normal state for all *n* observation periods;

– Class II - the object is in an abnormal state for all *n* observation periods;

– Class III - the object was in a normal state at the beginning of the observations, and in time period *k*, $1<k\leqslant n$, it went into an abnormal state.

As a result of the completed stages of the data preparation method, there are *N*•*n* MFLs.

The methodology of the analysis consists of the following stages.

Stage 1. Search for patterns in the MFL values.

A search is carried out for patterns characteristic only of class I objects in all *n* periods and class III objects in the periods included in the range (*0; k*). It is important to understand that these patterns must be absent in the MFL values of class II objects in all *n* periods and class III objects in the periods included in the range [*k; n*]. A pattern is considered to be the correspondence of certain variables to a certain factor in the MFL under consideration according to the search conditions described above. There may be no patterns between the values of Class II objects in all *n* periods and Class III objects in periods within the range [*k; n*], since the causes that caused the abnormal condition may be of different nature.

The block diagram of the described method is shown in Fig. 1. In the block diagram: 1 - comparative analysis of factor models, 2 - analysis of the dynamics of changes in factor loading models, crossed out arrows - absence of patterns.

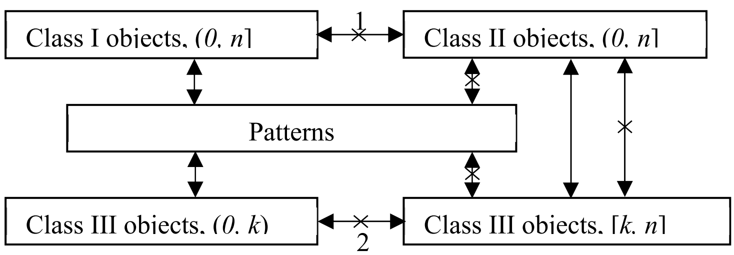

Fig. 1. Block diagram of the methodology for analyzing the dynamics of changes in factor loading models.

The search for patterns is carried out as follows:

- in each MFL for each variable, the maximum value by modulus is selected;
- each selected value corresponds to a certain factor;
- the presence/absence of correspondences of the selected values to the factors for all MFLs is checked.

Class I and II objects are necessary for the correct identification of patterns.

Stage 2. Analysis of the dynamics of changes in the MFN of class III objects.

The essence of the analysis at this stage is to fix the changes found at stage 1 of the patterns in the MFN of Class III objects. The found changes are analyzed to identify the fact of the transition of Class III objects from a normal state to an abnormal one.

4 Numerical experiment

The experience of conducting factor analysis for many areas of activity is currently presented quite widely. Including for solving production tasks during testing of electrical and radio components of space technology. To analyze the dynamics of changes in factor load models, we will conduct a numerical experiment based on the available data on the functioning of the low-orbit communication satellites.

Let's consider the application of factor analysis to test the hypothesis proposed in the introduction.

4.1 Hypothesis statement and methods of its verification

To test the hypothesis, data were taken from four spacecraft for two consecutive periods (conditional spacecraft numbers No. 01, 02, 03, 04). Of these, spacecraft No. 01 operated without any issues during the first period, and its technical condition became abnormal during the second period. Spacecraft No. 02 operated with issues during both periods. For both spacecraft, the issues affect the work for the intended purpose. The other two spacecraft (No. 03, 04) have no issues with their operation (they are in normal condition). The hypothesis is that the factor models calculated for different periods of spacecraft operation should differ depending on the change in its technical condition. The basis for testing the hypothesis is a change in the technical condition of spacecraft No. 01. Data on spacecraft No. 02, 03, 04 are taken as information necessary for conducting an analysis and checking the validity of the proposed hypothesis.

Hypothesis testing methodology:

- calculation of MFL for four spacecraft for two available periods (8 matrices);
- search for patterns characteristic of MFL spacecraft No. 03, 04 and for spacecraft No. 01 in the first period;
- the patterns found above should be absent in MFL spacecraft No. 02 and MFL spacecraft No. 01 in the second period;
- analysis of the dynamics of changes in patterns in MFL spacecraft No. 01 for two periods.

4.2 Description of input data

The variables selected as input data describe the state of different spacecraft systems and influences on the spacecraft.

Input data groups:
a) telemetry information from the spacecraft;
b) data from information resources about solar activity;
c) statistical information on the completion of the target task;
d) ballistic information;
d) other information about spacecraft.

To test the hypothesis, data were taken for each spacecraft for two separate periods lasting one year. Each observation is one day of the spacecraft operation. As a result, for each spacecraft we got 730 observations (two years) on 13 variables.

4.3 Data preparation for the experiment

Data preparation is carried out in accordance with section 2 of this paper. Some of the input data have a time scale of observations of 1 day. It is not possible to reduce the time scale for these data in our case. In this regard, we select this scale as optimal. Another part of the data has a smaller time scale. To conduct the experiment, we transform the time scale of these data to the optimal one.

Normalization of the input characteristics was carried out using the standard deviation (1).

The significance of the initial correlation matrices was verified using the Bartlett-Wilkes sphericity criteria and the adequacy of the Kaiser sample. The application of the Bartlett-Wilkes sphericity criterion confirmed the importance of correlation matrices. The value of the Kaiser sample adequacy criterion for all spacecraft ranged from 0.61 to 0.78, which indicates an acceptable and satisfactory adequacy of the factor model application.

Applying both criteria for a preliminary assessment of the number of allocated factors, set out in section 2, paragraph 6, we obtained an equal number of allocated factors for each spacecraft. For each spacecraft, only 4 factors will be considered.

According to Section 2 of this article, the principal component method was chosen for factor analysis of data describing the state of different spacecraft systems and external influences on the spacecraft. Varimax was chosen as the rotation method, since no significant difference in the results of factor analysis was observed when using other methods. Varimax maximizes the total variance of the squares of the loadings of common factors for each variable. Factor models in the numerical experiment are presented in the form of MFL.

4.4 The results of the experimental verification of the hypothesis

The MFL analysis, carried out in accordance with the methodology from section 3, allows us to explain the influence of factors on the spacecraft. The selected maximum modulus values of the variables interpret the factors as follows:

1) Thermal influence of the Sun. Thermal processes inside the spacecraft depend on the duration of the spacecraft's stay in illuminated or in shadowed areas of the orbit.

2) The influence of the power supply system state. It can be interpreted as the workload of the power supply system.

3) Electromagnetic and radiation influence of the Sun. This factor is not related to other variables and is the same for all MFLs, which means that these variables do not affect the performance of the spacecraft. This is explained by the fact that the electronic component base of the spacecraft used has high inherent reliability.

4) A factor that partially combines the influence of other variables.

If the first three factors are expressed quite stably, then the fourth factor requires detailed study for the analysis of the dynamics of changes in factor load models. In the process of searching, patterns were found related to variables from the groups of "statistical information" and "other information". It was concluded that abnormal situations in devices without issues do not have a great impact on their functioning.

Special attention is paid to changes in the factor loads of the variables for spacecraft No. 01. In the first period, the influence of the power supply system is explained by the fourth factor. For all other spacecrafts in all periods and for spacecraft No. 01 in the second period, the influence of the power supply system is explained by the second factor. Such a deviation of the matrix in the first period for spacecraft No. 01 from the other MFLs is probably due to the fact that spacecraft No. 01 was in a state of transition from normal to abnormal.

The fourth factor has a different interpretation for spacecraft with and without issues. Presumably, this factor can be interpreted as the influence of abnormal situations on the functioning of devices with issues. And for spacecrafts without issues, this factor combines the influence of the spacecraft operating time with traffic.

Other changes in the factor loads of variables that do not fit into the overall picture are presumably related to errors in the collection and insufficient statistical data.

As a result of the numerical experiment and the analysis carried out, it can be concluded that the factor models calculated for different periods of operation of the spacecraft for their intended purpose differ depending on changes in its technical condition. Which confirms the validity of the proposed hypothesis.

5 Conclusion

By applying the methods of factor analysis to predicting the functioning of spacecraft, it was possible to reduce a large number of input variables to a smaller number of independent quantities – factors. As a result, it became possible to identify key factors and interpret them in an accessible way and to present the influence of factors affecting the spacecraft in the form of MFL.

During the numerical experiment, the proposed hypothesis was developed and confirmed that the factor models calculated for different periods of spacecraft operation for the intended purpose differ depending on the change in its technical condition. In the process of implementing the data preparation methodology, 13 input variables were represented by four factors. As a result of interpreting the obtained factors, four groups were identified, each of which has its own impact on the functioning of the spacecraft. Further analysis of changes in the values in these groups revealed statistically stable patterns associated with changes in the technical condition of the spacecraft by periods.

The analysis of the dynamics of changes in factor load models was carried out by an expert method. This method has its drawbacks, such as subjectivity and the probability of error. The use of machine learning methods to search for patterns will allow more accurate interpretation of the results obtained and simplify the practical application of the methodology.

In practice, it is not always possible to assess the degree of influence of operational factors on the functioning of the spacecraft. Finding hidden relationships makes it possible to simplify complex systems and improve data interpretation. The proposed hypothesis helps to understand the structure of the relationships between the key factors influencing the object of study and its technical condition.

The proposed approach makes it possible to predict failures during spacecraft operation and the stability of space systems as a whole. Namely, by applying the described methods, it is possible to determine in advance which objects from the group under consideration are in the risk zone and are close to the moment of transition to an abnormal state. This allows for pre-planning the replenishment of the orbital group, carrying out restoration work, and making adjustments when developing future products. In addition, the proposed methods can be used not only in the space industry, but also for analyzing any complex systems in such industries as economics, medicine, agriculture, sociology, etc.

The work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2022-1121).

References

1. A. Büyul, P. Zeffel. *SPSS: the art of information processing. Analysis of statistical data and restoration of hidden patterns* (OOO "DiaSoftUP", St. Petersburg, 2005)
2. A. Nasledov, *IBM SPSS Statistics 20 and AMOS: professional statistical data analysis*. (Piter,St. Petersburg, 2013)
3. V.A. Ostreikovsky, *Reliability Theory* (Higher. school, Moscow, 2003)
4. N.S. Abramov, Software systems: theory and applications **6:2(25),** 85-99 (2015).
5. A.V. Donskov, D.A. Makhalov, A.V. Kuimov, Information and space **4** 82-90 (2022)
6. N.L. Rezova, L.A. Kazakovtsev, G. Sh. Shkaberina, M.I. Tsepkova, Control systems and information technologies, **88,** 2 (2022)
7. S.M. Mirak, *Neural network-based fault diagnosis of satellites formation flight: A thesis in The Department of Electrical and Computer Engineering* (Concordia University, Montreal, Canada, 2013)
8. D. Gras, *Data Science from Scratch* (O'Reilly Media, Inc., Sebastopol, USA, 2019)
9. A.N. Antamoshkin, I.S. Masich, Vestnik of SibGAU **5(57)** 20-25 (2014)
10. K. Iberla, *Factor analysis* (Statistics, Moscow, 1980)

11. Mazurov V.D. Factor analysis and the meaning of factors as a function of the meanings of attributes, in: *Russian regions in the focus of change: materials of the XI Int. Conf.,* 17-19 November 2016, (Yekaterinburg, Russia, 2016)
12. D. Lawley, A. Maxwell, *Factor analysis as a statistical method* (Butterworths, London, 1971)